

Electricity price forecasting in European Day Ahead Markets: a greedy consideration of market integration

TIES VAN DER HEIJDEN^{1,2,3}, (Student member, IEEE), JESUS LAGO⁴, PETER PALENSKY¹, (SENIOR MEMBER, IEEE), AND EDO ABRAHAM²

¹Department of Electrical Sustainable Energy, Delft University of Technology, Delft, the Netherlands

²Department of Water Resources Engineering, Delft University of Technology, Delft, the Netherlands

³HKV Consultants, Delft, the Netherlands

⁴Amazon, Amsterdam, The Netherlands

Corresponding author: Ties van der Heijden (e-mail: T.J.T.vanderHeijden-2@TUDelft.nl).

This work was supported by: TKI Watertechnologie.

Jesus Lago contributed to this work as an outside activity and not as part of his role at Amazon.

ABSTRACT In this manuscript we explore European feature importance in Day Ahead Market (DAM) price forecasting models, and show that model performance can deteriorate when too many features are included due to over-fitting. We propose a greedy algorithm to search over candidate countries for European features to be used in a DAM price forecasting model, that can be applied to several regression and machine learning modelling methodologies. We apply the algorithm to build price forecasting models for the Dutch market, using candidate countries selected through an integrated analysis based on open-source European electricity market data. Feature importance is visualised using an Auto Regressive and Random Forest model. We explain these results using cross-border flow and DAM price data. Two types of models (LEAR and the Deep Neural Network) are considered for the DAM price forecasting with and without European features. We show that in the Dutch case, taking European market integration into account improves the Mean Absolute Error (MAE) of the best performing DAM price forecasting model by 3.1%, the relative MAE (rMAE) by 3.85%, and the Symmetrical Mean Absolute Percentage Error (sMAPE) by 0.31 p.p., compared to the best forecasting model without European features. Through statistical testing we show that European features improve the accuracy of the forecasts with statistical significance.

INDEX TERMS Electricity price forecasting, European market integration, day ahead markets, deep neural networks, deep learning, feature analysis.

NOMENCLATURE

DR	Demand Response.	fARX	model using regression trees [1].
DAM	Day Ahead Market.	LEAR	The full-ARX model, as proposed in [2].
IDM	Intraday Market.		Lasso Estimated Auto Regressive, the fARX estimated with the LASSO as proposed in [3].
ENTSO-E	European Network for Transmission System Operators in Europe.	MLP	MultiLayer Perceptron.
EPF	Electricity Price Forecasting.	DNN	Deep Neural Network.
ML	Machine Learning.	SM-DNN	Single-market DNN, a DNN forecasting 24 hourly DAM prices of a single market.
ARX	Linear Auto Regressive model with Exogenous variables.	MM-DNN	Multi-market DNN, a DNN forecasting 24 hourly DAM prices of multiple markets simultaneously.
LASSO	Least Absolute Shrinkage and Selection Operator, a linear model estimation algorithm that includes regularisation.	TPE	Tree-structured parzen estimator, Bayesian hyperparameter optimisation algorithm [4].
RF	Random Forest, an ensemble regression	DM-test	Diebold-Mariano test, a statistical test to

compare accuracy between forecast time-series.

I. INTRODUCTION

DEMAND RESPONSE (DR) is an energy management technique where energy consumers are incentivised to shift their energy use in time [5], [6]. In market based DR, the incentive is economic through variable pricing. The electricity price reflects the availability of (renewable) energy through scarcity of a product. Low prices correspond with an energy surplus and low marginal cost of energy, where high prices correspond with energy scarcity and high marginal cost of energy. The introduction of spot markets allows for a different electricity price in every hour, 30-minutes and even 15-minutes. In Europe, the *Day Ahead Market* (DAM) is the main spot market for trading electricity. On the DAM, energy is traded in hourly blocks with corresponding prices. The DAM is well studied in the context of *Electricity Price Forecasting* (EPF) [7].

The European grid is well connected and the inter-connectivity of countries is expected to increase in the future. In 2002, members of the EU vowed to have 10% of their generation capacity in cross-border transmission capacity [8]. A future "European Supergrid" could even connect the European grid to North Africa [9]–[11]. Inter-connectivity of the European grid will help reach greater efficiencies, improve resilience to climate change and will enhance energy flexibility [12]. The development of European infrastructure will improve internal market efficiencies, enhance security of supply and enable *Renewable Energy Sources* (RES) market penetration [8]. Inter-connectivity enables cross-border electricity trading, opening up national electricity markets to foreign demand. These facts already make a strong case for the consideration of European market integration in a DAM price forecasting model. French market features (e.g. load, prices, generation), for example, have been shown to be more important than Belgian features when forecasting the Belgian DAM prices [13]. However, the analysis was limited to one neighbouring country and only a Deep Neural Network (DNN) was applied. Similarly, a study [14] has shown that including *Energy Exchange Austria's* (EXAA) prices as features improves forecast accuracy in all other EU DAM markets, especially in the Dutch and French DAM forecasting models. However, the connection between the Dutch DAM and the EXAA is also shown to be getting weaker over time. Possibly due to the further integration of the Dutch market with the APX UK market. In another study [15], price forecasts of other European markets are used as exogenous input for an Italian DAM forecasting algorithm, significantly improving performance. The effect of European market integration is not only seen in the DAM, but also in the *Intraday Market* (IDM) [16]. In general, the effect of European market integration is understudied in the context of EPF.

Machine Learning (ML) techniques have been shown to be effective at forecasting electricity prices, both in the DAM [7]

and the IDM [17], [18]. Specifically, the *MultiLayer Perceptron* (MLP) has been successfully applied to forecast DAM prices in Spain and Pennsylvania-New Jersey-Maryland [19], [20]. The more complex structured *Deep Neural Network* (DNN) was successfully applied in Belgium [21], Nord-Pool markets, Germany, France and Pennsylvania-New Jersey-Maryland [22].

Many other methods have been applied to forecast DAM prices [7]. According to an EPF benchmark study on the Belgian market, a two-layer DNN generally outperforms both statistical and other ML models, given the same features [21]. The *Lasso Estimated Auto Regressive* (LEAR) [22], or *fARX-Lasso* [3], is the best performing non-ML method and should be considered as a benchmark [22], especially since ML methods have been shown to not always outperform statistical models [21].

The performance of the different modelling approaches varies over different markets [22], but a recent case-study on the Dutch market is missing in literature. Many statistical methods rely on the calibration of linear relationships. While they can still be powerful modelling approaches, they might not perform well with high-resolution data like hourly prices with high volatility [21]. Price volatility in electricity markets can be pronounced due to the continuous need for balancing supply and demand [23]. It can differ significantly per month, and is subject to external price drivers like energy demand, demand elasticity, congestion, (renewable) energy generation, fuel prices, currency exchange rates and interconnected electricity markets [13], [23]–[25]. With an increasing renewable energy penetration in electricity markets, price volatility is expected to increase due to the intermittent nature of renewables, in the absence of sufficient energy storage [25]. However, effective policy, cooperation between TSO's and intraday trading can presumably prevent a significant increase in price volatility on the DAM [26]. It is possible that as renewable energy penetration increases, price volatility would increase and ML methods would increasingly outperform statistical methods.

In this manuscript we perform an EU wide, data-based analysis of European market feature (e.g. price and load) importance in DAM price forecasting models of European bidding zones. Open-source data from the ENTSO-E transparency platform [27] was used exclusively. Yearly cross-border flows are analysed, after which we apply a *Least Absolute Shrinkage and Selection Operator* (LASSO) [28], [29] estimated Auto Regressive model, and a *Random Forest* [1] model to identify European feature importance (Section IV). The results are used to select candidate European countries whose features are to be considered in a Dutch DAM forecasting model. We perform a benchmark using only Dutch features (Section V-B). Then, we propose a greedy approach to searches through candidate countries for the best performing combination of features for a DAM forecasting model (Section V-A). We apply two different types of models in the search for market integration features: the LEAR [3], [22] and the DNN. The LEAR is a linear regression model

estimated with the LASSO. The DNN will be applied in two different configurations: the *single-market DNN* [21] (SM-DNN) and the *multi-market DNN* [13] (MM-DNN). The SM-DNN [21] is a neural network that can forecast all 24 hourly DAM prices of a bidding zone simultaneously. The MM-DNN is a neural network that forecasts the 24 hourly prices of multiple bidding zones simultaneously. Temporal variations in relative model forecasting performance will be analysed (Section V-D) using univariate *Diebold-Mariano-tests* (DM-tests) on the hourly forecasts, in combination with Kernel Density Estimates of the daily DAM prices per hour of the day.

II. ELECTRICITY PRICE FORECASTING METHODS

In order to test whether our proposed methodology for including market integration features in a Dutch price forecasting model leads to a performance increase, several benchmarking models are proposed. Using a *naive* forecast as general benchmark, we compare its performance with 4 other Dutch price forecasting models without European features. The *full AutoRegressive* (fAR) and the *fAR with Exogenous variables* (fARX) [2], [3] are compared with the *Lasso Estimated AutoRegressive* (LEAR) [3], [22], consisting of the LASSO- or EN-estimated fARX. In this study we considered both the LASSO- and EN-estimated LEAR, and selected the best performing model based on a preliminary analysis. Since the *Deep Neural Network* (DNN) has recently been shown to be the best performing model in an EPF benchmark for the Belgian market [21], we include it in the benchmark for a Dutch price forecasting model.

The best performing models without European features will be used in the greedy algorithm in order to quantify their performance increase due to the inclusion of European features. The DNN will be included since it has already successfully been applied to improve the accuracy of the Belgian price forecast using French market features [13]. However, we consider two configurations of the DNN. First, we apply the SM-DNN with European features. The SM-DNN is used to forecast the 24 hourly Dutch DAM prices only, while European features are included in the input. Second, we apply and the MM-DNN, where the 24 hourly DAM prices of multiple bidding zones are forecast in the same model [13]. The MM-DNN is depicted in Figure 1, where the model is applied to forecast the Dutch and N other bidding zones' hourly DAM prices. We evaluate all models on their forecast accuracy of the Dutch DAM prices only.

A. LASSO ESTIMATED AUTO REGRESSIVE MODEL

The *Lasso Estimated Auto Regressive* (LEAR) is a fARX-model estimated using the LASSO (Appendix A1) [3], [22]. For DAM forecasting considered here, the fARX is given by

the following formula:

$$\begin{aligned}
 p_{d,h} = & \sum_{i=1}^{24} (\beta_{h,i} p_{d-1,i} + \beta_{h,i+24} p_{d-2,i} + \beta_{h,i+48} p_{d-3,i}) \\
 & + \beta_{h,73} p_{d-7,h} + \sum_{j=1}^3 (\beta_{h,j} p_{d-j}^{\min} + \beta_{h,j} p_{d-j}^{\max} + \beta_{h,j} p_{d-j}^{\text{avg}}) \\
 & + \beta_{h,83} z_{d,h} + \beta_{h,84} z_{d-1,h} + \beta_{h,85} z_{d-7,h} + \sum_{k=1}^7 \beta_{h,85+k} D_k \\
 & + \sum_{k=1}^7 \beta_{h,92+k} D_k z_{d,h} + \sum_{k=1}^7 \beta_{h,99+k} D_k p_{d-1,h} + \varepsilon_{d,h},
 \end{aligned} \tag{1}$$

where $p_{d,h}$ is the price on day d and hour h , β are the weights or coefficients, p_d^{\min} , p_d^{\max} and p_d^{avg} are the minimum, maximum and average price of day d , $z_{d,h}$ is the logarithm of the load for historic data, and the load forecast of day d , D_k are weekday dummies for all days of the week where holidays are all 0, and ε is the model error. The model can be written more compactly as

$$p_{d,h} = \sum_{i=1}^n \beta_{h,i} X_{d,h,i} + \varepsilon_{d,h}, \tag{2}$$

where $X_{d,h,i}$ are the $n = 106$ regressors or features in Equation (1).

The model is trained by solving the least-squares estimate of the model, with either an l_1 -norm for the LASSO-estimation or an $l_1 l_2$ -norm for the EN-estimation. With sufficiently large penalty term some weights will become (near) zero [30], effectively performing variable selection. More information on the LASSO- or EN- estimation can be found in Appendix A1.

B. DEEP NEURAL NETWORK

A DNN is a feed forward *Artificial Neural Network* (ANN) that is trained using back propagation. It contains a minimum of 3 layers: an input layer, an output layer and at least 1 hidden layers. Following the standard notation in EPFs, we will use the name MLP for ANNs with 1 hidden layer and DNN for ANNs with 1 or more. Although the definition of what is "deep" might be debatable, it is out of the scope of this paper and we simply follow the current standard terminology used in literature [13], [21], [22].

The DNN is trained with the *Adaptive Moment Estimation* (ADAM) [31]. Figure 1 shows an MM-DNN with 2 hidden layers, where n_f is the number of input features, n_1 and n_2 are the number of nodes in hidden layer 1 and 2, respectively. The output layer (P) consists of 24 nodes for each country included in the forecast, one for each hour of the day. In the case of Figure 1, the DNN is used to forecast Dutch DAM prices together with a variable number (N) of other bidding zones. In the case of a DNN with one hidden layer, the output layer is fully connected to the first hidden layer.

Features are optimised as hyperparameters similarly to [13], where a single binary choice-option for country load

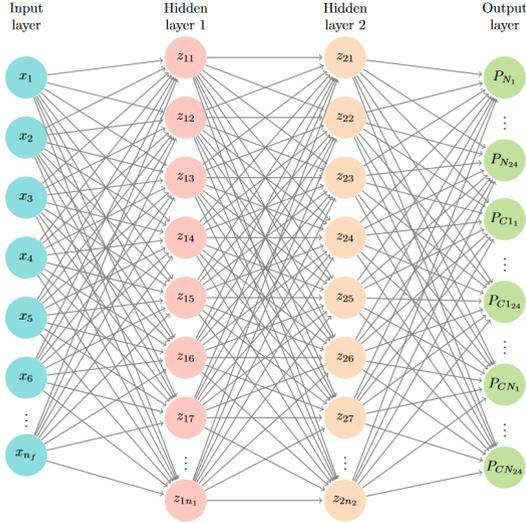


FIGURE 1: Multi-Market Deep Neural Network with two hidden layers applied to forecast Dutch DAM prices simultaneously with N other bidding zones' prices. If $N = 0$, the DNN forecasts a single market's prices, the Dutch prices, only. If $N = 1$, the dual-market DNN as proposed in [13] results.

features is added to the search space. In our case these are the historic load, and load forecast of the countries considered in the forecast, which are added or removed through a single binary choice-option per country. Historic prices are always included as model features.

To optimise hyperparameters and features, the *Tree-structured Parzen Estimator* [4] (TPE) algorithm is used. The TPE is a Bayesian *Sequential Model Based Optimisation* (SMBO) methods, which can efficiently be applied to optimise hyperparameters [32], [33]. The method consists of a surrogate model that is built using Bayes rule, describing the probability of model performance as a function of the hyperparameters

$$p(y|x) = \frac{p(x|y) * p(y)}{p(x)}, \quad (3)$$

where y is model performance and x is a hyperparameter instantiation.

The surrogate model is used to estimate model performance as a function of features and hyperparameters. In the TPE algorithm, $p(x|y)$ is defined such that

$$p(x|y) = \begin{cases} l(x) & \text{if } y < y^* \\ h(x) & \text{if } y \geq y^* \end{cases}, \quad (4)$$

where $l(x)$ is the density formed by observations $x^{(i)}$ such that the corresponding loss, $f(x^{(i)})$, is less than a threshold (y^*). $h(x)$ is the density formed by the remaining observations. The TPE algorithm samples candidates from $l(x)$, after which $\frac{l(x)}{h(x)}$ is evaluated for each sample in order to suggest the candidate with the highest expected improvement.

C. DIEBOLD-MARIANO TEST

The DM-test is a statistical measure to compare accuracy of two forecast timeseries. Given a target timeseries (y_t) and two forecasts ($\hat{y}_{1,t}$, $\hat{y}_{2,t}$), the error timeseries ($e_{i,t}$) are calculated (Equation 5a). In this manuscript we apply the Mean Absolute Error (MAE) (Equation (5b)) as loss function (\mathcal{L}). Based on this, a loss differential timeseries (δ_t) is calculated (Equation 5c), which is then used to perform the one-sided DM-test.

$$e_{i,t} = \hat{y}_{i,t} - y_{i,t}, \quad i = 1, 2 \quad (5a)$$

$$\mathcal{L}_{i,t} = \|e_{i,t}\|_1, \quad (5b)$$

$$\delta_t = \mathcal{L}_{1,t} - \mathcal{L}_{2,t}, \quad (5c)$$

The *one-sided DM-test* tests the null hypothesis (H_0) that forecast timeserie 1 is more accurate than or equally accurate as forecast timeserie 2. The alternative hypothesis (H_1) is that forecast 2 is more accurate than forecast 1. In the context of DAM forecasting, the DM-test can be applied to either the full DAM timeseries (e.g. multivariate DM-test) or to evaluate forecasting performance for each hour separately (e.g. univariate DM-test).

$$H_0 : E(\delta_t) \leq 0, \quad (6a)$$

$$H_1 : E(\delta_t) > 0 \quad (6b)$$

D. GREEDY ALGORITHM FOR EUROPEAN FEATURE SEARCH

In order to include EU market integration in a bidding zone's DAM forecasting model, we propose a greedy algorithm as defined in Algorithm 1. We apply the algorithm to prevent over-fitting on the large amount of available data by selecting only the relevant features. Starting with a candidate set of features (e.g. Θ_1 contains only Dutch features), a model (\mathcal{M}_i) is trained and its hyperparameters are optimised. The model is evaluated to test its predictive performance (p_i) on the bidding zone of interest's prices, in our case the Dutch bidding zone's, and added to a set (\mathcal{P}_i) together with model performance (p_i) and features (Θ_i) of the current iteration (i). After the first iteration, all candidate countries (Φ) are consecutively added to the latest feature set (Θ_i), to be used in the price forecasting model. The model is trained, its hyperparameters are optimised, and its forecasting performance on the bidding zone of interest (\mathcal{P}_j) is evaluated and added to set (\mathcal{P}). After all candidate countries are evaluated, the best performing model's features are selected from the set (\mathcal{P}). Those features are added to the feature set (Θ_i), while they are removed from the candidate feature set (Φ). The procedure will then be repeated until either performance stops increasing, or all candidate features are added to the model.

The candidate set of features (Φ) was selected using the data analysis described in Section IV, where all countries that result from at least one of the analyses (cross-border flow, ARX or RF) are selected. The candidate set of country features used for the Dutch market are summarised in Table 1.

Our proposed greedy algorithm can be applied to any model that allows for a large amount of features, and assists with preventing over-fitting. In this manuscript we independently apply the algorithm to three different modelling approaches (Section V-C). We perform both hyperparameter optimisation and feature selection on the validation set of the Dutch DAM prices. Finally, all models are evaluated on the test set to evaluate the increase in model performance. No model selection is performed on the test set.

Algorithm 1: GreedyFeatureSearch(Θ_1, Φ)

```

1  $i \leftarrow 1$ 
2  $\mathcal{P} \leftarrow \emptyset$ 
3  $p \leftarrow \emptyset$ 
4  $\mathcal{M} \leftarrow \emptyset$ 
5  $\mathcal{M}_i \leftarrow \text{TrainModel}(\Theta_i)$ 
6  $p_i \leftarrow \text{EvaluateModel}(\mathcal{M}_i)$ 
7  $\mathcal{P}_i \leftarrow (\Theta_i, p_i, \mathcal{M}_i)$ 
  while  $(p_i \leq p_{i-1} \vee i > I) \wedge |\Phi| > 0$  do
    // execute algorithm while
    // performance improves, or until all
    // country features are included in
    // the model
8    $P \leftarrow \emptyset$ 
9    $M \leftarrow \emptyset$ 
    for  $j = \{1, \dots, |\Phi|\}$  do
      // train and evaluate the  $i^{\text{th}}$  model
      // by adding a single country's
      // ( $j$ ) features at a time
10     $M_j \leftarrow \text{TrainModel}(\{\Theta_i \cup \Phi_j\})$ 
11     $P_j \leftarrow \text{EvaluateModel}(M_j)$ 
    end
12     $j \leftarrow \arg \min_j (P)$ 
13     $\Theta_{i+1} \leftarrow \{\Theta_i \cup \Phi_j\}$ 
    // add best performing country's
    // features ( $\Phi_j$ ) to  $i^{\text{th}}$  model
14     $\Phi \leftarrow \{\Phi \setminus \Phi_j\}$ 
15     $i \leftarrow i + 1$ 
16     $p_i \leftarrow P_j$ 
17     $\mathcal{M}_i \leftarrow M_j$ 
18     $\mathcal{P}_i \leftarrow (\Theta_i, p_i, \mathcal{M}_i)$ 
  end
19 return  $\mathcal{P}$ 

```

III. DATA AND TOOLS

The data used for this study is taken from the ENTSO-E transparency platform [27]. Historic DAM price-series, historic loads and historic day ahead load forecasts are used. The data was split in train-, validation- and test-sets, as depicted in Figure 2. Data from 2015-2019 was used, with the year 2019 being used as test data. In order to be able to include the most recent information in training, 2017 was taken as validation data instead of 2018. A single split limits the data leakage that occurs in the first week of the 2018, where data

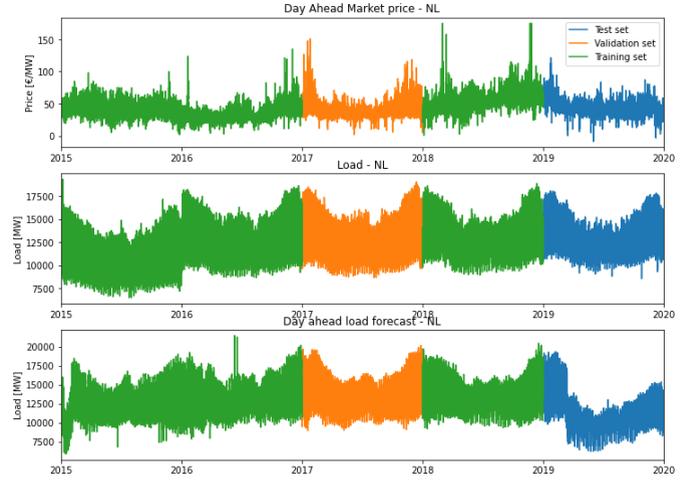


FIGURE 2: Dutch historic price, load and load-forecast with train, validation and test splits indicated for model training.

from the validation set is used in the lagged features. There still is no data leakage from the test set.

The LEAR models are build and trained, and the features were Z-score standardised using the scikit-learn python package [34]. The DNN was made using Tensorflow [35] and trained with the ADAM optimiser [31]. The TPE SMBO algorithm is implemented using the Python Hyperopt package [4]. The DM-tests are performed using the python EPF-toolbox [22].

IV. EUROPEAN FEATURE IMPORTANCE ANALYSIS

In order to identify possible European features that influence the Dutch market, a EU-wide analysis is performed on cross-border flow data and DAM prices.

European cross-border flow data [27] was analysed to identify trading patterns between countries. Figure 3 shows the cross-border flows between countries whose data is available through the ENTSO-E transparency platform. The data shows that France is the largest net exporter of the EU, while Italy is the largest net importer of the EU. The Nordic markets (FI, SE, DK and NO) seem to be a relatively independent block, mostly trading amongst each other.

For each bidding zone, a LASSO-estimated ARX-model is trained and its weights are analysed. The model is given by the following formula:

$$p_{d,h} = \sum_{i=1}^N \beta_i p_{i,d-1,h} + \sum_{j=1}^M \beta_{j+N} z_{j,d,h} + \varepsilon_{d,h}, \quad (7)$$

where N is the number of bidding zones, M the number of countries, $p_{i,d,h}$ the price of bidding zone i and $z_{j,d,h}$ the load forecast of country j , on day d and hour h , β are the weights or coefficients, and $\varepsilon_{d,h}$ the model error. The analysis allows us to gain insight in the relative importance of European features with respect to domestic features. The analysis fits the general EPF modelling approach, where historic prices and the load (forecast) are used to forecast next

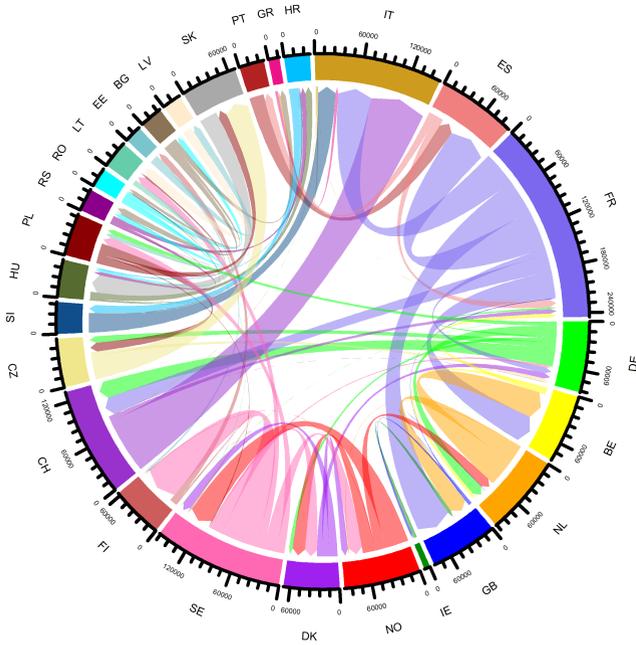


FIGURE 3: Cross-border flows between European countries, units in [MWh/year] based on 2018 data [27]

day’s price. Figure 4a shows the normalised weights resulting from the ARX-models. The incoming arrow width indicates the relative weight of transmitting country features for the receiving country’s price forecasting model. For countries containing multiple bidding zones, the weights of the bidding zones are summed. Weights are masked at a minimum of 5% normalised importance.

Since the LASSO estimated ARX is a linear model, it is possible that non-linear relationships are not well reflected in the analysis. To overcome this limitation, a RF-model is applied in similar fashion. The RF is trained to forecast the 24 hourly DAM prices of a bidding zone simultaneously, using the same features as the ARX-model. The RF is able to capture non-linear relationships, and can similarly show relative features importance. More information on the RF model can be found in Appendix A2. Figure 4b shows the normalised feature importance resulting from the RF-models. The RF seems to give a higher importance to domestic market features compared to the ARX.

Both the ARX- and RF-analyses show that Norwegian market features seem to perform well in many other European price forecasting models, while Norwegian models perform well using only Norwegian features. Compared to Sweden, which exports more in an absolute sense, Norwegian features perform well in a large number of forecasting models of other European markets. This could be due to the amount of storage present in the Norwegian system, possibly allowing them to export more dynamically. It would enable them to export at times the merit-order curve in other bidding zones is relatively high and steep, expressing more influence on price

settlement. Another explanation could be the relative large transmission losses that come with the overseas cables that connect Norway to many other European markets, making high foreign prices necessary to make up for the transmission losses. Figure 5 shows the correlation matrix of Norwegian export and the DAM Prices of the importing markets. The figure shows that Norwegian exports correlates positively with the importing market’s DAM prices. The correlation is the strongest in Denmark and Sweden, but is positive for all neighbouring markets. The positive correlation between Norwegian export and Swedish prices is as positive as to other connected market prices. This indicates that the difference between overland- and overseas-connected markets, and as such the possible effect of transmission losses on export behaviour, can’t be observed to have an effect on the correlation between Norwegian export and the neighbouring price.

The Spanish and Portuguese markets seem to function as an independent block from the other EU markets, with French features having low importance in their price models. This could partially be explained by the Pyrenees separating them, which makes overland transmission cables relatively expensive [36]. This would result in the relatively low degree of interconnection of Spain with other EU member states [37].

Even though France is the largest net exporter of electricity, its feature importance in neighbouring market models is relatively low. One explanation could be the large share of nuclear power in France, making its export possibilities relatively inflexible. This makes it possible that the information in French features is captured in the historic prices of those markets, and less information from France is needed to quantify its effect on price settlement. Figure 6 shows the correlation matrix of French electricity export and the importing market’s DAM prices. The figure shows that the correlation between electricity export and DAM prices is weak, and in some cases negative. This can similarly explain the relative low importance of Italian features in other markets’ models, regardless of its high net import. Figure 7 shows the correlation matrix of Italian import and the transmitting market’s DAM prices. Again, only weak correlations are observed. A constant Italian demand in foreign markets (e.g. Slovenian, Swiss and French) could be captured in the historic prices of those markets, and less Italian information is needed to quantify this effect. Although cumulatively, Italian feature importance is relatively high in both the ARX- and RF-models.

Germany and Switzerland trade a significant amount of electricity, but their features are not selected each other’s price forecasting model. Germany has a large renewable energy share, making a constant trading behaviour less likely. However, it could be that a part of the export is from intraday trading. If that is the case, cross-border trading would have a smaller effect on DAM prices.

Both the ARX- (Figure 4a) and the RF-analyses (Figure 4b) show there is no straightforward relationship between

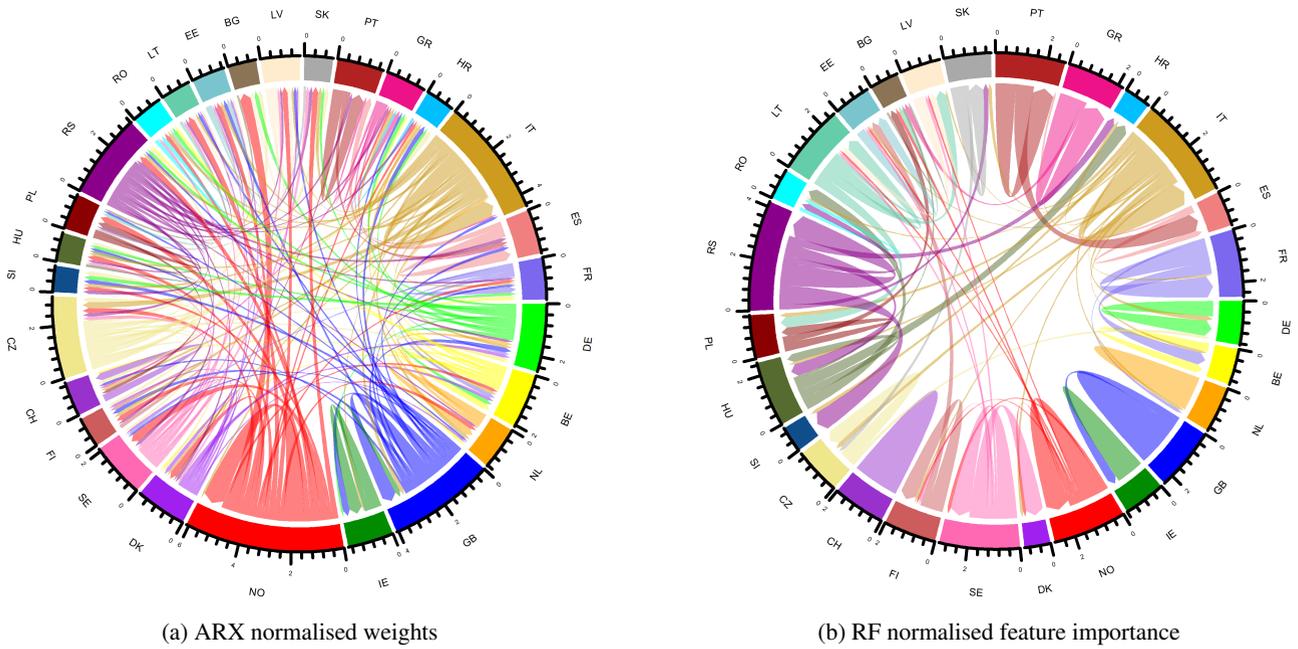


FIGURE 4: Normalised weights for the ARX model (a) and normalised feature importance for the RF model (b) forecasting the DAM prices of day d , using the prices of day $d - 1$ and the load forecast of day d of all European bidding zones and countries. The weights of the load and separate bidding zone's prices are summed for each country.

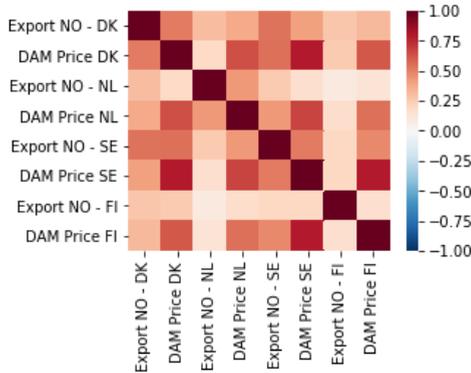


FIGURE 5: Correlation matrix of Norwegian electricity export and the DAM prices of importing markets. Based on 2018 data.

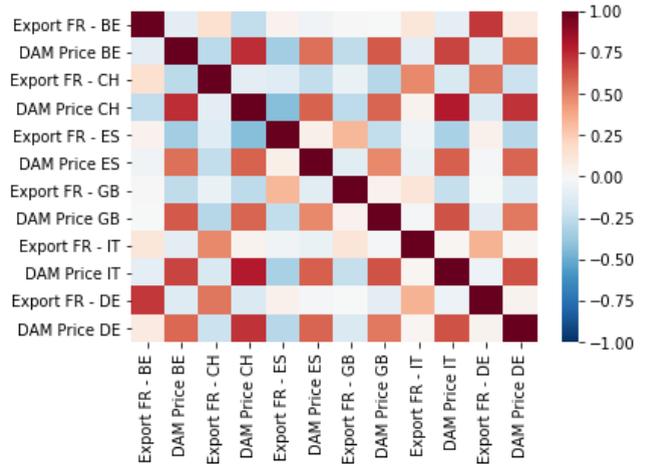


FIGURE 6: Correlation matrix of French electricity export and the DAM prices of importing markets. Based on 2018 data.

cross-border flows and feature importance in a foreign price forecasting model. Norwegian features have high importance in many markets, while French feature importance is limited. France exports a lot more electricity than Norway, but their inter-connectivity, means of generation and as such their inherent flexibility are vastly different. Which can also be seen in the correlation matrix of their export and the importing markets' DAM prices. Italian features show up in many ARX-models (Figure 4a), but with low importance compared to other markets regardless of its large net import. A possible explanation is the ability of a bidding zone to time its import and export with advantageous prices. However, more research is necessary to conclude this definitely.

For the Dutch price model, Table 1 summarises the results of the analysis perform in this section. It stands out that directly connected markets (BE, FR, NO and GB) arise from the ARX- and RF-analyses. The ARX- and RF-models identify second-order market effects through Danish (DK), French (FR) and Italian (IT) features. German features are considered due to the relatively large amount of trading between the two markets. For the Belgian, British and German price forecasting models, Dutch features seems to contain information on their prices as well. However, for the French,

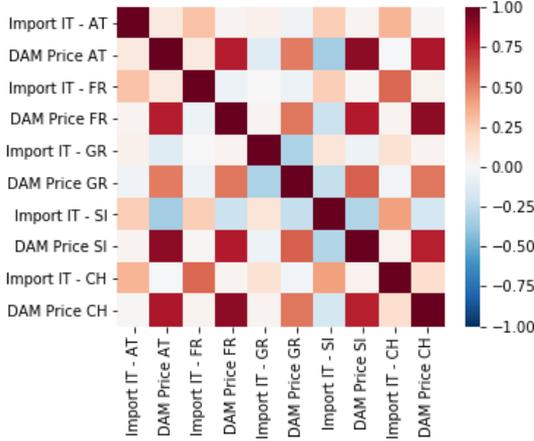


FIGURE 7: Correlation matrix of Italian electricity import and the DAM prices of exporting markets. Based on 2018 data.

Italian, Danish and Norwegian models, Dutch features are not of high importance.

V. DUTCH DAM PRICE FORECASTING WITH EU MARKET INTEGRATION

To search for the best combination of European features for several Dutch price forecasting models, we apply the greedy algorithm (Algorithm 1) to three modelling approaches independently. The algorithm will be applied to search over the previously defined candidate countries (Φ). Table 1 shows the summarised results for the European feature candidate selection, applied to the Dutch market. For countries containing multiple bidding zones (NO, DK, IT), a single bidding zone (NO-2, DK-1, IT-NORD) is selected for the remaining analysis.

Cross-border flows	LASSO	Random Forest
Germany	Norway	Italy
Norway	Italy	France
Belgium	Belgium	
Great Britain	Denmark	
France	Great Britain	

TABLE 1: Candidate countries to be included in the search for Dutch price forecasting features, based on cross-border flows, ARX-, and RF feature importance. A country is selected as candidate when it shows up in at least one of the analyses.

A. MODEL TRAINING

For model training of the LEAR and DNN, two slightly different approaches are taken. In the LEAR models, all fARX-variables are considered possible features for the candidate countries. The penalty-factors (λ , ρ) are optimised on the validation set. For the DNN, a feature- and hyperparameter search is performed using the TPE algorithm (Section II-B). The search space (Φ) that we have defined for the DNN's is

shown in Table 2. The best performing models, evaluated on the validation set, are selected for all modelling approaches.

Hyperparameter	Variable type	Search space
Number of layers	Integer	[1,2]
Nodes per layer	Integer	Layer 1: 50 - 450 Layer 2: 50 - 250
Dropout rate	Continuous	0 - 0.5
Batch size	Integer	$7^1 - 7^4$
Regularisation (l_2 -norm)	Continuous	$1e^{-5} - 5e^{-2}$
Batch normalisation	Binary	[False, True]
Random seed	Integer	1 - 300
Load features	Binary	[False, True]

TABLE 2: TPE search space for the DNN models. Load features are added as a hyperparameter for every country that is considered in the forecast.

B. DUTCH BENCHMARK

In order to be able to say whether the inclusion of European features improves the performance of the Dutch forecast, a benchmark is performed using the Naive, fAR, fARX [2], [3], LEAR [3], [22] and DNN [21] models without European features. For the fAR, fARX and LEAR, a single model is trained for each hour of the day. While the DNN is applied to forecast the 24 prices simultaneously.

The results in Table 3 show that the LEAR sets a tough benchmark to beat, outperforming the DNN. Figure 8a shows the results of the multivariate DM-test. The test confirms that the LEAR model's forecast is significantly more accurate than all other forecasts using only Dutch features, while the DNN's forecast is significantly more accurate than all others' but the LEAR's.

	Naive	fAR	fARX	LEAR	DNN
MAE	€5.68	€4.78	€4.63	€4.40	€4.52
rMAE	1	0.85	0.82	0.78	0.80
sMAPE	14.45%	11.91%	11.70%	11.00%	11.36%

TABLE 3: Performance on the test set (2019) of the benchmark models for Dutch DAM price forecasting.

C. GREEDY SEARCH FOR EUROPEAN FEATURES

Algorithm 1 is applied to search for the best feature combination in a Dutch DAM forecasting algorithm. The LEAR and DNN models are applied in the search for European features, where the DNN is both applied to forecast the Dutch price only (SM-DNN), and as a multi-market forecasting model (MM-DNN). All models are evaluated for their ability to forecast Dutch DAM prices only. Table 4 shows the forecast metrics of the LEAR models, showing an increase in accuracy as features from more countries are included. Model performance stopped increasing after 5 added countries (GB, DK, NO, FR and DE). The largest improvement can be found after the second and fourth iteration, when Danish and French features are included in the model. The uneven increase in performance could mean that there are system effects in price settlement, which are hard to quantify using only two markets. The LEAR_{S_{all}} model using all European features

shows similar performance to the LEAR with features from set $S_{L,1}$. It still has improved performance over the LEAR model with only Dutch features. This can be explained by the strong regularisation ability of the LASSO, effectively enforcing sparsity in the solution. However, there is still some degree of over-fitting resulting in a lower loss than most models with less features. Figure 8b shows the results of the multivariate DM-test of the forecasts. It shows that as features from more countries are included in the model, the performance increases significantly. This leaves the LEAR with features from set $S_{L,5}$ as statistically significant best performing LEAR-model.

Table 5 shows the results from the European feature search applied to the SM-DNN. No significant increase can be seen after the first iteration, where British features are added to the model. After three iterations, test performance increased with respect to the Dutch-only model. However, the fourth iteration resulted in the model with the lowest test loss. Interestingly, the same countries except Germany are selected as in the LEAR model in different order. Model performance decreased after the second iteration, while the validation loss decreased. This could indicate that relevant market dynamics can change depending on the year. The inclusion of Norway lead to the largest loss decrease, which can possibly be explained by Norway’s high feature importance found in Section IV. Figure 8c shows the results of the multivariate DM-test between the forecasts of the SM-DNN’s. It shows little statistical significance in improved forecast accuracy. The model with features from set $S_{D,4}$ does show some significant performance increase. The SM-DNN with features from set $S_{D,4}$ has the most convincing p-values for improved forecast accuracy compared to models with features from sets NL , $S_{D,1}$ and $S_{D,3}$. In general, the forecast error decreased by including more European features. Even though not all improvements are statistically significant, the final model including features from set $S_{D,4}$ is arguably better than the DNN with only Dutch features. The SM-DNN $_{S_{all}}$ model with all European features shows a large performance decrease, even compared with the SM-DNN using only Dutch features. This indicates that even though the degree of l_2 regularisation is taken as a hyperparameter, the model is hard to train with this amount of features.

Table 6 shows the performance of the models generated in the European feature search for the MM-DNN. The model is trained by minimizing the MAE of the price forecasts over all countries included in the model. Naturally, trade-offs are made between, for example, Dutch and French price forecasting accuracy. This could both lead to a performance decrease due to trade-offs, or to a performance increase by preventing over-fitting. The greedy algorithm select the same markets in the same order as the single market DNN, but stops one iteration sooner.

The MM-DNN with features from $S_{D,2}$ and $S_{D,3}$ outperform the single market DNN with the same features. Indicating that forecasting prices of multiple markets could indeed lead to better single-market performance. Forecast accuracy

generally improved with the inclusion of more European features. The MM-DNN with countries from set $S_{D,3}$ is significantly better than all other MM-DNN models and the DNN using only Dutch features. Although the significance of the improvement from $S_{D,2}$ to $S_{D,3}$ is debatable. The MM-DNN $_{S_{all}}$ model with all European features shows a large performance decrease. The rMAE larger than 1 shows the forecast performance is lower than the naive model. The larger performance decrease in the MM-DNN could be explained by a trade-off between target accuracy’s. The loss function while training the model combines the losses for all bidding zones, which could result in a trade-off between price forecasting performance in the Dutch and other bidding zones.

Table 7 shows the countries whose features are added in every iteration and for all applied modelling approaches. Table 8 shows the improvements of the applied modelling approaches when European market integration features are included in the model. All approaches show improvements with an increasing number of market integration features. For the Netherlands, the first countries whose features are included by the greedy algorithm are Great Britain and Denmark for all modelling approaches. French features are included in all modelling approaches, and in some cases Norwegian and German features are included as well. The third country added by the greedy algorithm differs between the LEAR and the DNN’s. In the LEAR, Norway is chosen as third external market, while in the DNN French features are included. This could indicate that the relationship between the Norwegian market and Dutch prices are more linearly approachable than the relationship between the Dutch price and French market. A possible explanation is that the effect of the dynamics between the French and British market on the Dutch price, are better captured in the DNN’s.

The largest improvements are found in the MM-DNN, but it doesn’t perform better than the LEAR. The large relative improvement of the MM-DNN could mean that market integration dynamics are better captured by non-linear relationships, resulting in a larger possible improvement in non-linear models in comparison to linear models.

The SM-DNN doesn’t perform as well as the MM-DNN, which could be explained by two factors: during training, it is possible that the MM-DNN’s were able to overcome local optima when different features are included. This would result in improvements in accuracy that are not due to the information in the data, but due to the training process. Another factor could be the prevention of over-fitting by generalizing tasks [13]. However, extensive optimisation of the DNN-type model performance was not within the scope of the research. Meaning that it is possible that other DNN configurations or other training methods could result in DNN models that outperform the LEAR.

D. TEMPORAL VARIATION IN PERFORMANCE

While the multivariate DM-tests give a good relative representation of forecast accuracy, electricity prices are known to

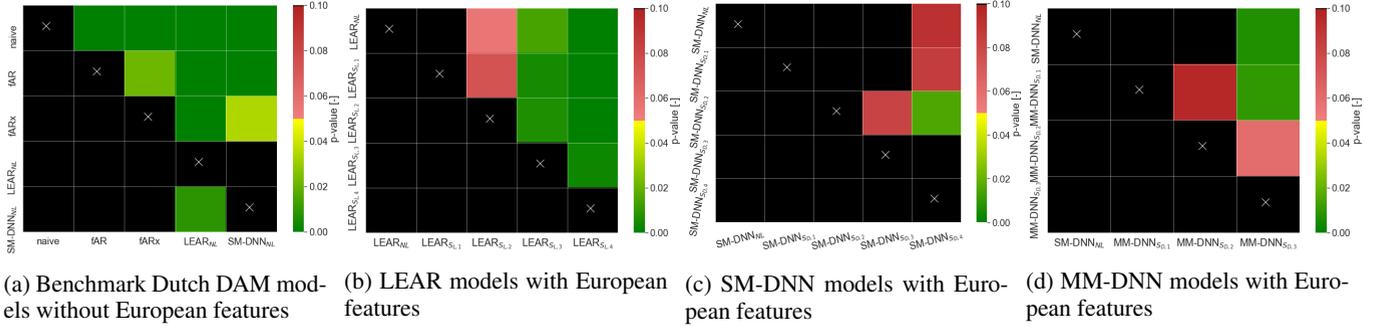


FIGURE 8: Multivariate DM-test results of the Dutch DAM benchmark models (a), the LEAR models with European features (b), the SM-DNN models with European features (c), and the MM-DNN models with European features (d). In the figure, the color indicates the p-value (green-yellow: $p \leq 0.05$, light red-dark red: $0.5 > p > 0.1$, black: $p \geq 0.1$) of the one-sided DM-test on whether the forecast of the model on the X-axis is more accurate than the forecast of the model on the Y-axis.

	NL	$S_{L,1}$	$S_{L,2}$	$S_{L,3}$	$S_{L,4}$	$S_{L,5}$	S_{all}
MAE	€4.40	€4.39	€4.35	€4.33	€4.29	€4.26	€4.38
rMAE	0.78	0.77	0.77	0.77	0.76	0.75	0.77
sMAPE	11.00%	11.02%	10.89%	10.85%	10.76%	10.69%	10.92%

TABLE 4: Dutch DAM LEAR models with European features. Notation: $S_{L,1} = \{NL, GB\}$, $S_{L,2} = \{S_{L,1}, DK\}$, $S_{L,3} = \{S_{L,2}, NO\}$, $S_{L,4} = \{S_{L,3}, FR\}$, $S_{L,5} = \{S_{L,4}, DE\}$, S_{all} includes all European features in the model.

	NL	$S_{D,1}$	$S_{D,2}$	$S_{D,3}$	$S_{D,4}$	S_{all}
MAE	€4.52	€4.52	€4.58	€4.49	€4.43	5.61
rMAE	0.80	0.80	0.81	0.79	0.78	0.99
sMAPE	11.36%	11.31%	11.47%	11.26%	11.08%	13.79%

TABLE 5: Single-market DNN forecasting models with European features. Notation: $S_{D,1} = \{NL, GB\}$, $S_{D,2} = \{S_{D,1}, DK\}$, $S_{D,3} = \{S_{D,2}, FR\}$, $S_{D,4} = \{S_{D,3}, NO\}$, S_{all} includes all European features in the model.

	NL	$S_{D,1}$	$S_{D,2}$	$S_{D,3}$	S_{all}
MAE	€4.52	€4.53	€4.45	€4.36	€6.29
rMAE	0.80	0.80	0.79	0.77	1.11
sMAPE	11.36%	11.38%	11.19%	10.95%	15.33%

TABLE 6: Multi-market DNN forecasting models, evaluated on Dutch price forecasting performance only. Notation: $S_{D,1} = \{NL, GB\}$, $S_{D,2} = \{S_{D,1}, DK\}$, $S_{D,3} = \{S_{D,2}, FR\}$, S_{all} includes all European features in the model.

be more volatile in some hours than others. Figure 9 shows the Kernel Density Estimates (see Appendix A3) of the Dutch DAM prices in August and October 2019. It can be seen that the DAM price behaves differently throughout the day, with low price volatility in the early morning and late evening, and more volatile prices during midday. Figure 10 shows the univariate DM-test of LEAR model with features from $S_{L,5}$ against the benchmark LEAR and DNN, $DNN_{S_{D,4}}$ and $MM-DNN_{S_{D,3}}$. It can be seen that the LEAR forecasts are more accurate than the DNN models' forecasts mostly at hours with low price volatility. Figure 11 shows the same DM-test but applied to model $MM-DNN_{S_{D,3}}$. It can be seen that the MM-DNN forecasts are more accurate than the LEAR type model forecasts at times with some degree of volatility,

	LEAR	SM-DNN	MM-DNN
Iteration 1	GB	GB	GB
Iteration 2	DK	DK	DK
Iteration 3	NO	FR	FR
Iteration 4	FR	NO	-
Iteration 5	DE	-	-

TABLE 7: Countries whose features are selected by the greedy algorithm for every iteration and per model type.

	LEAR	DNN	MM-DNN
Δ MAE [%]	3.1	2.0	3.5
Δ rMAE [%]	3.85	2.5	3.75
Δ sMAPE [p.p.]	0.31	0.29	0.41

TABLE 8: Performance increase of the applied modelling approaches due to the inclusion of market integration features.

but not when it is too high. These months have been chosen for the sake of simplicity, the results can also be seen in other months. Renewable energy generation forecasts are not included as features, but could assist with forecasting highly volatile prices. The analysis indicates that as prices are becoming more volatile, stochastically trained non-linear models start outperforming mathematically optimised and regularised linear models. This makes the DNN modelling approach worth considering for the future Dutch DAM forecast, as it is possible that electricity prices will become more volatile in the future [25].

VI. CONCLUSION

In this manuscript, we propose a method for searching optimal combinations of European features in a Day Ahead Market (DAM) price forecasting models. By using the Dutch market as a case study, we show that taking European market integration into account in regression and machine learning

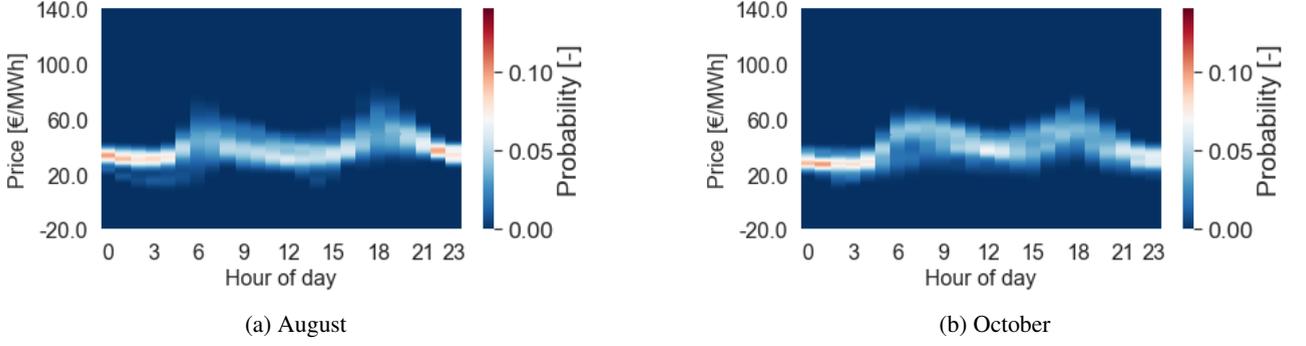


FIGURE 9: Kernel density estimates of Dutch DAM prices in August (a) and October (b) 2019. The X-axis shows the hours of the day, the Y-axis shows the price in [€/MWh], and the color shows the probability of occurrence [-].

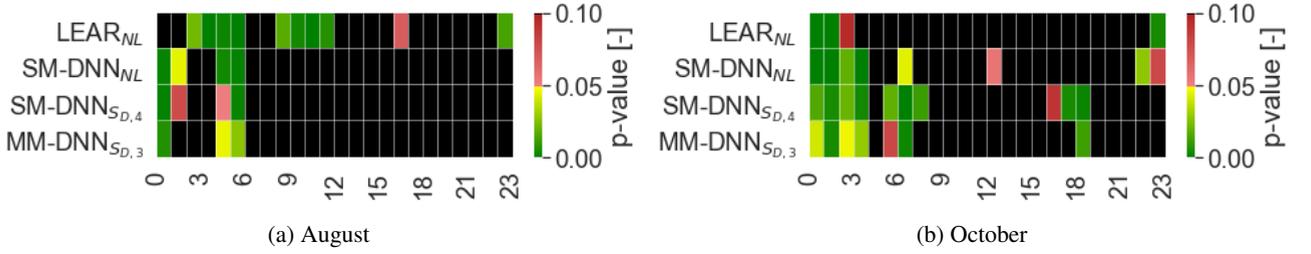


FIGURE 10: Univariate DM-test of LEAR_{SL,5} model's hourly forecasts in August (a) and October (b), being more accurate than the hourly forecasts of the models on the Y-axis.

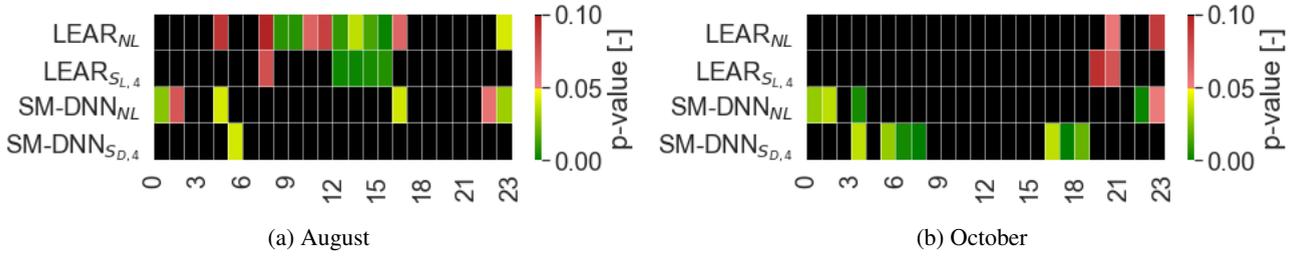


FIGURE 11: Univariate DM-test of MM-DNN_{S_D,3} model's hourly forecasts in August (a) and October (b), being more accurate than the hourly forecasts of the models on the Y-axis.

models can improve the forecasting performance of the best performing DAM models. We identify and visualize European feature importance, showing that flexibility in the energy system might affect a country's feature importance. This is confirmed by the correlation between electricity export and DAM prices. We have proposed a greedy algorithm to search for European features in a DAM price forecasting model, using candidate countries resulting from an analysis on before-mentioned feature importances. The proposed greedy algorithm improved forecasting performance of all proposed modelling approaches with statistical significance. The algorithm was applied to the Dutch market, using three different modelling approaches: the LEAR [3], [22], the Single-Market Deep Neural Network (SM-DNN) [21] and the Multi-Market DNN (MM-DNN) [13]. A benchmark was performed using the LEAR and DNN including only Dutch features. The algorithm was able to identify features while

preventing over-fitting on the data, which did happen in the models using all European features. The algorithm first included British and Danish market features, after which either Norwegian or French were included in the model, differing between the LEAR and the DNN's. The LEAR model, which is known to have good regularisation properties, performs best when features from 5 other countries are included in the model. The SM-DNN model performed best with features from 4 other countries, and the MM-DNN performed best with features from 3 other countries. The largest improvement was found in the MM-DNN, which can be explained by either the training process (as discussed in Section V-C) or the non-linear nature of market dynamics. However, in the Dutch case the LEAR model with market integration features is the overall best performing model. Temporal variations in model performance are shown, and a relationship between price volatility and relative model

performance was found. The LEAR generally performs better than the DNN during hours with low volatility, while the DNN-type models show some improved performance over the LEAR in hours with high volatility. This could indicate that as prices are expected to become more volatile in the future [25], non-linear models could outperform linear models. The LEAR model performed relatively well using all European features, which can be explained by its strong regularisation properties. The DNN type models using all European features showed a decrease in performance, also with respect to using only Dutch features. The MM-DNN showed the largest performance decrease, which could be due to the multi-market loss function. This could result in trade-offs between price forecasting performance on the Dutch markets and other bidding zones. In future work, other modelling approaches could be used in the proposed greedy algorithm. Also, different lengths of the training data could be used, other hyperparameter configurations of the DNN could be set, and model ensembles could be made, in order to optimise forecasting performance. More features could be used, like renewable energy generation forecasts. Weights could be added to the MM-DNN loss function to optimize Dutch price forecasting performance. Other years of data can be used to train and test the models, allowing the investigation of temporal changes in market dynamics. Also, the analysis could be applied to different European markets to test the sensitivity of a certain market's model accuracy to market integration features. This would also allow for an extended analysis on the relative performance of linear and non-linear models under different scenarios of price volatility.

REFERENCES

- [1] L. Breiman, "ST4_Method_Random_Forest," *Machine Learning*, vol. 45, no. 1, pp. 5–32, 2001.
- [2] A. Misiorek, S. Trueck, and R. Weron, "Point and interval forecasting of spot electricity prices: Linear vs. non-linear time series models," *Studies in Nonlinear Dynamics and Econometrics*, vol. 10, no. 3, 2006.
- [3] B. Uniejewski, J. Nowotarski, and R. Weron, "Automated variable selection and shrinkage for day-ahead electricity price forecasting," *Energies*, vol. 9, no. 8, 2016.
- [4] J. Bergstra, D. Yamins, and D. Cox, "Making a Science of Model Search: Hyperparameter Optimization in Hundreds of Dimensions for Vision Architectures," *Proceedings of the 30th International Conference on Machine Learning*, Atlanta, Georgia, USA, 2013, 2013.
- [5] European Commission, "2050 low-carbon economy roadmap," European Commission, Tech. Rep., 2016. [Online]. Available: https://ec.europa.eu/clima/policies/strategies/2050_en
- [6] A. R. Jordehi, "Optimisation of demand response in electric power systems, a review," *Renewable and Sustainable Energy Reviews*, vol. 103, no. July 2018, pp. 308–319, 2019. [Online]. Available: <https://doi.org/10.1016/j.rser.2018.12.054>
- [7] R. Weron, "Electricity price forecasting: A review of the state-of-the-art with a look into the future," *International Journal of Forecasting*, vol. 30, no. 4, pp. 1030–1081, 2014. [Online]. Available: <http://dx.doi.org/10.1016/j.ijforecast.2014.08.008>
- [8] European Commission, "Achieving the 10% electricity interconnection target," European Commission, Tech. Rep. 9, 2015. [Online]. Available: https://ec.europa.eu/energy/sites/ener/files/publication/FOR_WEB_energy_union_interconnections_EN-1.pdf
- [9] W. Platzer and Fraunhofer ISE, "SUPERGRID – Approach for the integration of renewable energy in Europe and North Africa," p. 38, 2016. [Online]. Available: <https://www.ise.fraunhofer.de/en/publications/veroeffentlichungen-pdf-dateien-en/studien-und-konzeptpapiere/study-supergrid-fraunhofer-ise.pdf>
- [10] House of Commons, London, "A European Supergrid," 2011. [Online]. Available: <http://www.publications.parliament.uk/pa/cm201314/cmselect/cmenergy/194/194we03.htm>
- [11] M. Benasla, D. Hess, T. Allaoui, M. Brahimi, and M. Denäi, "The transition towards a sustainable energy system in Europe: What role can North Africa's solar resources play?" *Energy Strategy Reviews*, vol. 24, no. August 2018, pp. 1–13, 2019. [Online]. Available: <https://doi.org/10.1016/j.esr.2019.01.007>
- [12] European Commission, "An energy policy for Europe," European Commission, Tech. Rep., 2007.
- [13] J. Lago, F. De Ridder, P. Vranx, and B. De Schutter, "Forecasting day-ahead electricity prices in Europe: The importance of considering market integration," *Applied Energy*, vol. 211, no. July 2017, pp. 890–903, 2018. [Online]. Available: <https://doi.org/10.1016/j.apenergy.2017.11.098>
- [14] F. Ziel, R. Steinert, and S. Husmann, "Forecasting day ahead electricity spot prices: The impact of the EXAA to other European electricity markets," *Energy Economics*, vol. 51, pp. 430–444, 2015. [Online]. Available: <http://dx.doi.org/10.1016/j.eneco.2015.08.005>
- [15] I. P. Panapakidis and A. S. Dagoumas, "Day-ahead electricity price forecasting via the application of artificial neural network based models," *Applied Energy*, vol. 172, pp. 132–151, 2016. [Online]. Available: <http://dx.doi.org/10.1016/j.apenergy.2016.03.089>
- [16] C. Kath, "Modeling intraday markets under the new advances of the cross-border Intraday Project (XBID): Evidence from the German intraday market," *Energies*, vol. 12, no. 22, pp. 1–35, 2019.
- [17] M. Narajewski and F. Ziel, "Ensemble Forecasting for Intraday Electricity Prices: Simulating Trajectories," *Applied Energy*, vol. 279, no. April, p. 115801, 2020. [Online]. Available: <http://arxiv.org/abs/2005.01365>
- [18] C. Monteiro, I. J. Ramirez-Rosado, L. A. Fernandez-Jimenez, and P. Conde, "Short-term price forecasting models based on artificial neural networks for intraday sessions in the Iberian electricity market," *Energies*, vol. 9, no. 9, 2016.
- [19] D. Keles, J. Scelle, F. Paraschiv, and W. Fichtner, "Extended forecast methods for day-ahead electricity spot prices applying artificial neural networks," *Applied Energy*, vol. 162, pp. 218–230, 2016.
- [20] P. M. Bento, J. A. Pombo, M. R. Calado, and S. J. Mariano, "A bat optimized neural network and wavelet transform approach for short-term price forecasting," *Applied Energy*, vol. 210, no. October 2017, pp. 88–97, 2018. [Online]. Available: <https://doi.org/10.1016/j.apenergy.2017.10.058>
- [21] J. Lago, F. De Ridder, and B. De Schutter, "Forecasting spot electricity prices: Deep learning approaches and empirical comparison of traditional algorithms," *Applied Energy*, vol. 221, no. January, pp. 386–405, 2018. [Online]. Available: <https://doi.org/10.1016/j.apenergy.2018.02.069>
- [22] J. Lago, G. Marcjasz, B. D. Schutter, and R. Weron, "Forecasting day-ahead electricity prices: A review of state-of-the-art algorithms, best practices and an open-access benchmark," *Renewable and Sustainable Energy Reviews*, 2020 (Under review).
- [23] M. Benini, M. Marracci, P. Pelacchi, and A. Venturini, "Day-ahead market price volatility analysis in deregulated electricity markets," *Proceedings of the IEEE Power Engineering Society Transmission and Distribution Conference*, vol. 3, no. SUMMER, pp. 1354–1359, 2002.
- [24] S. Goodarzi, H. N. Perera, and D. Bunn, "The impact of renewable energy forecast errors on imbalance volumes and electricity spot prices," *Energy Policy*, vol. 134, no. June, p. 110827, 2019. [Online]. Available: <https://doi.org/10.1016/j.enpol.2019.06.035>
- [25] S. Mosquera-López and A. Nursimulu, "Drivers of electricity price dynamics: Comparative analysis of spot and futures markets," *Energy Policy*, vol. 126, no. October 2018, pp. 76–87, 2019. [Online]. Available: <https://doi.org/10.1016/j.enpol.2018.11.020>
- [26] A. Khoshrou, A. B. Dorsman, and E. J. Pauwels, "The evolution of electricity price on the German day-ahead market before and after the energy switch," *Renewable Energy*, vol. 134, pp. 1–13, 2019.
- [27] ENTSO-E, "Entso-e transparency platform," 2018. [Online]. Available: <https://transparency.entsoe.eu/>
- [28] R. Tibshirani, "Regression Shrinkage and Selection Via the Lasso," *Journal of the Royal Statistical Society: Series B (Methodological)*, vol. 58, no. 1, pp. 267–288, 1996.
- [29] S. L. Kukreja, J. Löfberg, and M. J. Brenner, "a Least Absolute Shrinkage and Selection Operator (Lasso) for Nonlinear System Identification," *IFAC Proceedings Volumes*, vol. 39, no. 1, pp. 814–819, 2006.
- [30] T. Hastie, R. Tibshirani, and M. Wainwright, *Statistical Learning with Sparsity: The Lasso and Generalizations*. CRC Press, 2016, vol. 84, no. 1.
- [31] D. Kingma and J. Ba, "Adam: A method for stochastic optimization," *International Conference on Learning Representations*, 12 2014.
- [32] F. Madrigal, C. Maurice, and F. Lerasle, "Hyper-parameter optimization tools comparison for multiple object tracking applications," *Machine Learning and Applications*, vol. 30, no. 2, pp. 269–289, 2019.
- [33] K. Eggenberger, F. Hutter, H. H. Hoos, and K. Leyton-Brown, "Efficient benchmarking of hyperparameter optimizers via surrogates," *Proceedings of the National Conference on Artificial Intelligence*, vol. 2, pp. 1114–1120, 2015.
- [34] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay, "Scikit-learn: Machine learning in Python," *Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, 2011.
- [35] M. Abadi, A. Agarwal, P. Barham, E. Brevdo, Z. Chen, C. Citro, G. S. Corrado, A. Davis, J. Dean, M. Devin, S. Ghemawat, I. Goodfellow, A. Harp, G. Irving, M. Isard, Y. Jia, R. Jozefowicz, L. Kaiser, M. Kudlur, J. Levenberg, D. Mané, R. Monga, S. Moore, D. Murray, C. Olah, M. Schuster, J. Shlens, B. Steiner, I. Sutskever, K. Talwar, P. Tucker, V. Vanhoucke, V. Vasudevan, F. Viégas, O. Vinyals, P. Warden, M. Wattenberg, M. Wicke, Y. Yu, and X. Zheng, "TensorFlow: Large-scale machine learning on heterogeneous systems," 2015, software available from tensorflow.org. [Online]. Available: <http://tensorflow.org/>
- [36] D. Elliott, "Emergence of European supergrids - Essay on strategy issues," *Energy Strategy Reviews*, vol. 1, no. 3, pp. 171–173, 2013. [Online]. Available: <http://dx.doi.org/10.1016/j.esr.2012.04.001>
- [37] J. Serrano González and C. Álvarez Alonso, "Industrial electricity prices in Spain: A discussion in the context of the European internal energy market," *Energy Policy*, vol. 148, no. October 2020, 2021.
- [38] D. Scott, *Multivariate Density Estimation: Theory, Practice, and Visualization*. John Wiley & Sons, New York, Chichester, 1992.
- [39] P. Virtanen, R. Gommers, T. E. Oliphant, M. Haberland, T. Reddy, D. Cournapeau, E. Burovski, P. Peterson, W. Weckesser, J. Bright, S. J. van der Walt, M. Brett, J. Wilson, K. J. Millman, N. Mayorov, A. R. J. Nelson, E. Jones, R. Kern, E. Larson, C. J. Carey, Í. Polat, Y. Feng, E. W. Moore, J. VanderPlas, D. Laxalde, J. Perktold, R. Cimrman, I. Henriksen, E. A. Quintero, C. R. Harris, A. M. Archibald, A. H. Ribeiro, F. Pedregosa, P. van Mulbregt, and SciPy 1.0 Contributors, "SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python," *Nature Methods*, vol. 17, pp. 261–272, 2020.



TIES VAN DER HEIJDEN (M'21) was born in Wageningen, the Netherlands in 1993. He received both his BSc and MSc degrees in Civil Engineering at Delft University of Technology, in 2019. In 2020 he started his PhD at Delft University of Technology, the departments of Water Resources Engineering and Electrical Sustainable Energy. His research focuses on the application and economic potential of Demand Response in open canal systems.

In 2019, he started working as Water and Energy consultant at HKV Consultants in Delft, while being active as a researcher at Delft University of Technology. His research interests include forecasting, electricity markets, European market integration, demand response and water system control.



PETER PALENSKY (M'03-SM'05) received his MSc. in Electrical Engineering in 1997 his PhD in 2001, and his Habilitation in 2015, all from the Vienna University of Technology, Austria. He co-founded a Envidatec, a German startup on energy management and analytics, and joined the Lawrence Berkeley National Laboratory, CA, as researcher, and the University of Pretoria, South Africa, in 2008. In 2009 he became appointed head of business unit on sustainable building technologies at the Austrian Institute of Technology (AIT), and later the first Principle Scientist for complex energy systems at the AIT. In 2014 he was appointed full Professor for intelligent electric power grids at TU Delft. He is active in international committees like ISO or CEN and serves an IEEE IES AdCom member-at-large in various functions for the IEEE. He is editor in chief of the IEEE Industrial Electronics Magazine, associate editor for several other IEEE publications, and regularly organizes IEEE conferences. His main research fields are energy automation networks, smart grids, and modeling intelligent energy systems.

gies at the Austrian Institute of Technology (AIT), and later the first Principle Scientist for complex energy systems at the AIT. In 2014 he was appointed full Professor for intelligent electric power grids at TU Delft. He is active in international committees like ISO or CEN and serves an IEEE IES AdCom member-at-large in various functions for the IEEE. He is editor in chief of the IEEE Industrial Electronics Magazine, associate editor for several other IEEE publications, and regularly organizes IEEE conferences. His main research fields are energy automation networks, smart grids, and modeling intelligent energy systems.



EDO ABRAHAM received an M.Eng degree in Electrical and Electronic Engineering and a Ph.D. degree in Control Engineering (Aeronautics) from Imperial College London, UK, in 2008 and 2013, respectively. During 2014-2016, he was a Research Associate in the Environmental and Water Resources Engineering Group, Imperial College London. Since 2016, he is an assistant Professor in Water & Control at TU Delft's Faculty of Civil Engineering and Geosciences.

His research focus is mainly in the application of mathematical optimisation, control and systems theory to advance water management and environmental engineering and their nexuses with energy and agricultural systems. His application areas span from open canal water systems, irrigation to space heating. Edo serves as an Associate Editor of the Springer journal Energy Systems.

A. ADDITIONAL METHODOLOGY DETAILS

In this Appendix we present an extended description of some of the methodologies used.

1) LASSO and Elastic-Net

The *Least Absolute Shrinkage and Selection Operator* (LASSO) [28] is a least squares linear regression model, where an additional l_1 -norm penalty is applied to the weights in order to regularise the model and enforce sparsity of the solution. The LASSO solves the optimisation problem

$$\min \frac{1}{2} \|z - \Phi\theta\|_2^2 + \lambda \|\theta\|_1, \quad (8)$$

where $z \in \mathbb{R}^{N \times 1}$ is a vector of outputs, $\Phi \in \mathbb{R}^{N \times p}$ is a matrix of features and $\theta \in \mathbb{R}^{p \times 1}$ is a vector of weights. The regularisation parameter $\lambda \in \mathbb{R}$ can be used to control the trade-off between sparsity of the solution and the approximation error.

The *Elastic-Net* (EN) is a linear regression model fit with both an l_1 and l_2 penalty norm on the weights of the model. During model selection, both the penalty gradient on the weights (λ) and the $l_1 l_2$ ratio (ρ) are optimised. A ρ of 0 would result in a Ridge Regression problem, while a ρ of 1 results in a LASSO. The objective function when fitting the EN is:

$$\min \frac{1}{2} \|z - \Phi\theta\|_2^2 + \lambda \rho \|\theta\|_1 + \frac{\lambda(1-\rho)}{2} \|\theta\|_2^2, \quad (9)$$

where variables have the same representation as in the LASSO objective (8), and ρ represents the $l_1 l_2$ ratio.

2) Random Forest

The *Random Forest* (RF) algorithm was developed by Breiman [1]. It is an ensemble classification or regression approach using random regression trees, hence the name Random Forest. The RF is a collection of tree-predictors whose output - in the case of regression - is an unweighted average of the outcomes of the separate trees:

$$\bar{h}(x) = \frac{1}{K} \sum_{k=1}^K h(x; \theta_k), \quad (10)$$

where \bar{h} is the averaged output of regression trees $h(x; \theta_k)$ with input x , feature space θ , subspace $\theta_k \subset \theta$ and K regression trees. The RF estimates a features's importance based on the amount of splits in the trees in comparison to the amount of samples the feature splits. It is also referred to as the 'Gini importance' or the 'mean decrease impurity' [34].

3) Kernel Density Estimation

A kernel density is a non-parametric method that can be used to estimate the probability density function (PDF) of a random variable. If (x_1, x_2, \dots, x_n) is a sample from the univariate distribution f , its kernel density estimator is then described by

$$\hat{f}_h(x) = \frac{1}{nh} \sum_{i=1}^n K\left(\frac{x - x_i}{h}\right), \quad (11)$$

where K is the kernel (in our case we have applied the Gaussian kernel function) and h is the bandwidth (in our case

this has been selected using Scott's method [38]). The KDE has been applied using the Scipy python package [39].